DIGITIZER INPGRAPH TO INPUT NUMERIC DATA INTO EXFOR LIBRARY

G.N. Pikulina, S.M. Taova,

October 22-25 2018, IAEA, Vienna, Austria

In the course of data compilation for the international library EXFOR it is required to provide the input of valid numerical nuclear reaction data. Very often such data are available from old journals where they are presented only in the form of plots. And it is impossible to obtain the numeric data from the authors of old articles. In this case digitization is the only way to get numeric data.

Generally, digitization is a conversion of an object, image (in the analogue form) into the form of a set of discrete values using special equipment, i.e. it is a transformation into a digital format suitable to be recorded on electron carriers. For example, a graphic presentation of experimental results is scanned in two dimensions to get a discrete image.

In our case the term «digitization» means getting of numeric data from an image obtained at original document scanning, at copying from a PDF-file or at capturing a part of the screen.

The task of getting table numeric values from image curves is not a particular one applicable only at preparing data for the EXFOR library. Below are given typical situations when digitization is required [1]:

– at working with literature if tables or analytical formulas are inaccessible and the data are available only in the form of plots that should be used in the calculations;

– for comparing measurement results and literary data in the form of plots or for comparing data from several literary sources where the plots are presented;

– at creating and completing databases on literary sources containing graphic information;

– for the tasks of inputting data basing on oscillograms obtained with the aid of a storage oscilloscope at measuring fast processes.

Basing on the statistics of May 2018 [3] more than 5700 descriptions of experiments (Entries) containing more than 34 thousand numeric data tables obtained from digitized images are presented in the EXFOR data base. This is about a quarter of all EXFOR Entries (total number of Entries is more than 22 000) and the same portion is related to the numeric data tables (total number of data tables is more than 154 000). So digitization widely used to get numeric data in the EXFOR format increases considerably the EXFOR content.

In the early 2000-s the specialists of the VNIIEF CNPD developed a digitizer InpGraph as a part of the EXFOR-Editor software package [3].

The basic advantage of InpGraph consisted in applying special mathematic treatment of digitized data to convert curve coordinates from the image coordinate scale to the experiment

coordinate scale. This feature allowed digitizing plots of poor quality and assisted in calculating the digitization error. Such possibilities provided the program competitiveness and acuteness. Among disadvantages one should mention that the digitizing algorithm required the strict order of operations and necessity of using several programs to get the result numeric values.

There was decided to update the InpGraph digitizer to give a user a modern easy-to-use interface and high-capacity mechanisms of image digitization and to exclude additional laborious operations. Simultaneously a new version should have used the proved mathematical procedures under the control of integrated software environment.

During the InpGraph program development field experience of the previous version and feedbacks of users were taken into account. Functional possibilities of digitizers free available in Internet (for example, GetData Graph Digitizer, Graph2Digit, G3Date etc. [4−6]) and programs developed by Japanese and Chinese Nuclear Reactions Data Centers (GSys, GDGraph [7,8]) were analyzed.

As a result the following tasks were solved by our specialists:

− the processing of images imported from different sources was provided: files of different graphic format, from clipboard and through capturing the selected screen area;

− the preprocessing of images was provided: skewing and rotation are available;

− the input of information according to the EXFOR rules was optimized;

− the EXFOR dictionaries were included;

− the digitization process was improved – i.e. the setting of axes in the automated mode procedure, the processing of curves with asymmetric errors were implemented;

− the calculation of quantization error was added.

A flexible program structure to provide the improvement and optimization of its functionality and to add new possibilities to the digitizing procedure was developed.

The InpGraph new version makes it possible to return and repeat any digitization step at any stage of image processing. The digitizing order is optional.

At the development of the InpGraph user interface the Wizard technology was implemented. Such helper function automatically presents step-by-step instructions and prompts for the user. The principle of minimum information input by the user was also applied: meaning selection from specified lists, context search for independent variables and data verification at the input stage were implemented.

The major task of the new interface is to maximally reduce the number of operations the user is to undertake and simplify the digitizing procedure itself. «The less clicks you are to make, the better is the program for a user».

InpGraph works under the control of Microsoft Windows operating system. Basic program functions are available from the InpGraph main window presented in Fig.1.

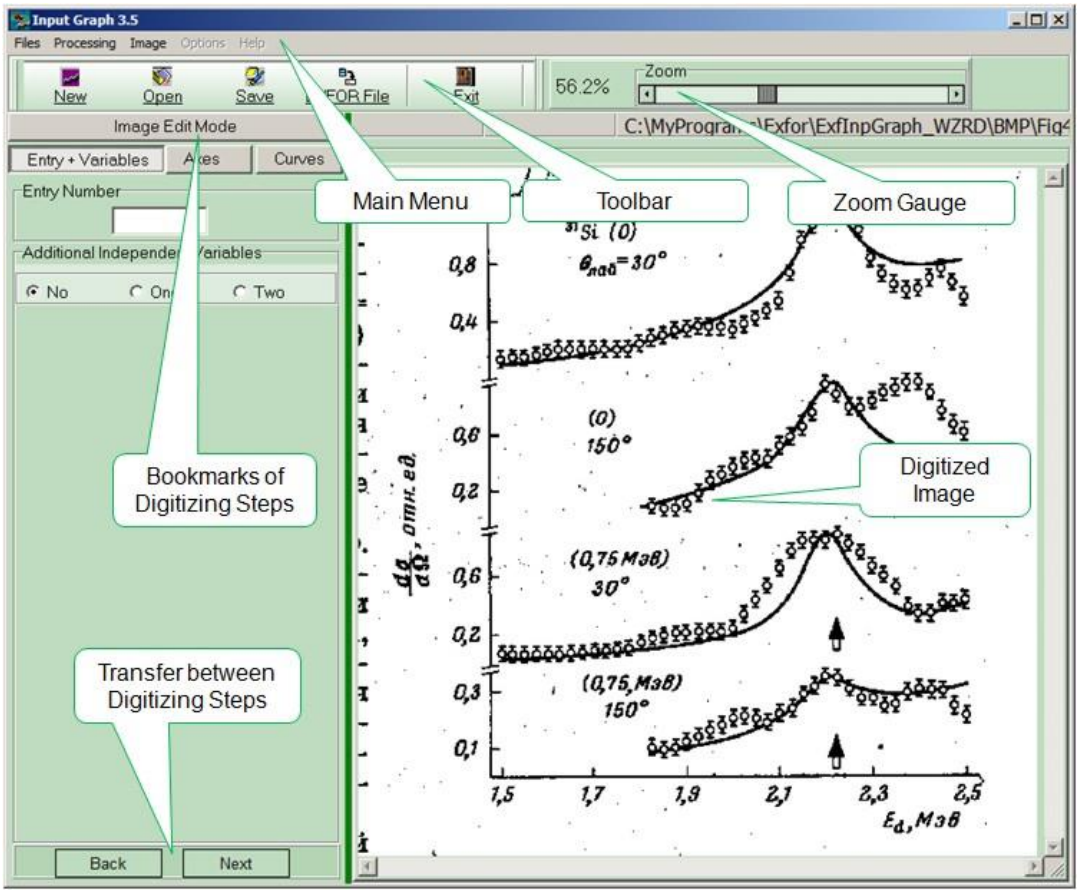

Fig. 1 – The InpGraph main window

Let us consider the digitizing process with the InpGraph program by a small example.

After loading the image to be digitized the Image Edit Mode bookmark is active. It contains the buttons for image preprocessing: rotation, skewing − these operations correct distortions occurring at scanning original images. The example of rotation procedure is presented in Fig 2.

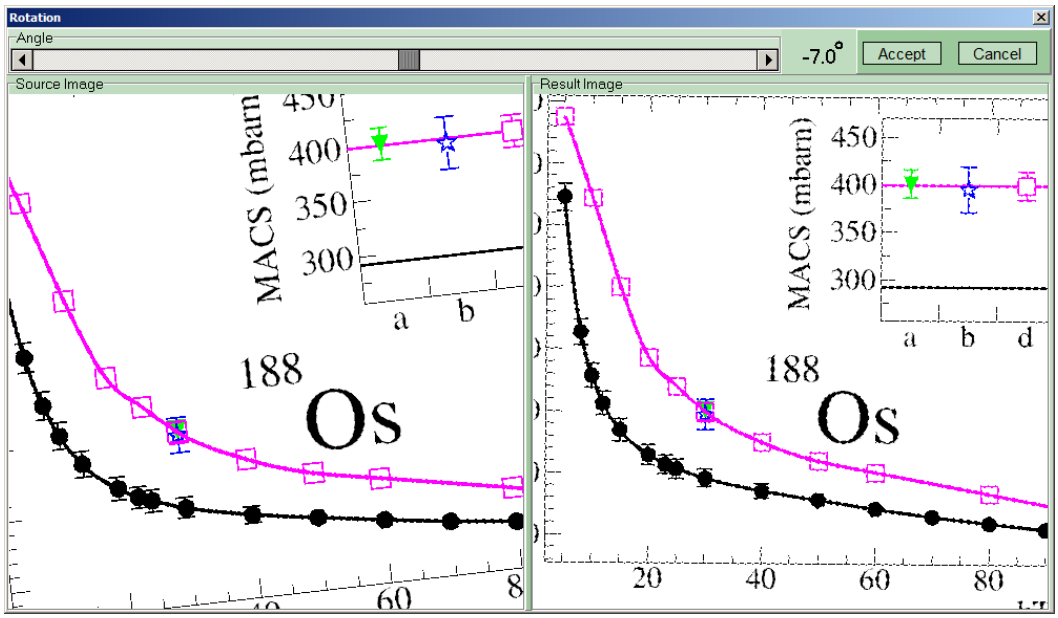

Fig. 2 – Example of image rotation

The next step is input of headings and units of additional independent variables from the EXFOR dictionaries (see Fig. 3). One may select only those values that are permitted by the EXFOR rules. There exists the possibility of context searching the necessary headings in the dictionaries using a combination of symbols.

Then, one has to use the Axes bookmark to input headings and units of axes and to set their direction, scale and scale ticks (see Fig. 4). The number of X- and Y-axes used is not limited. The axes points can be set manually or in an automated mode. Checking of axes linearity and accuracy of their setting is performed at the input stage.

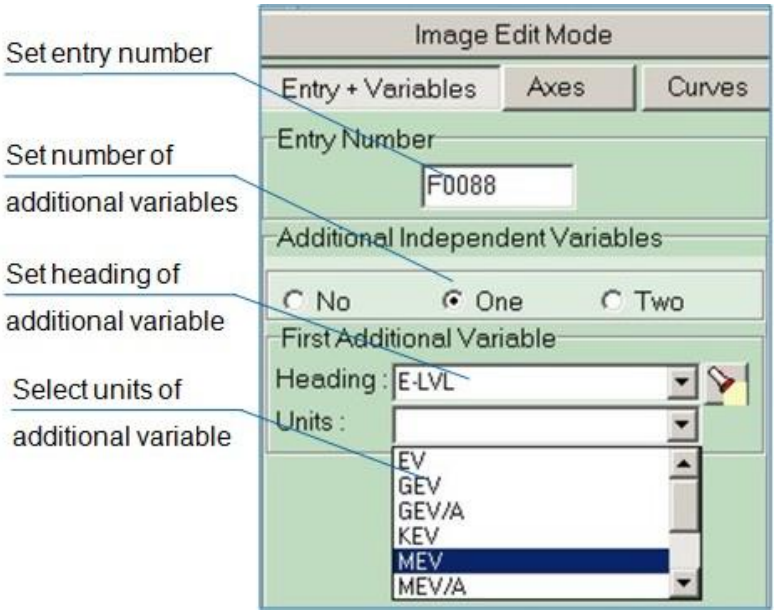

Fig. 3 − Input of names and dimensions of additional independent variables

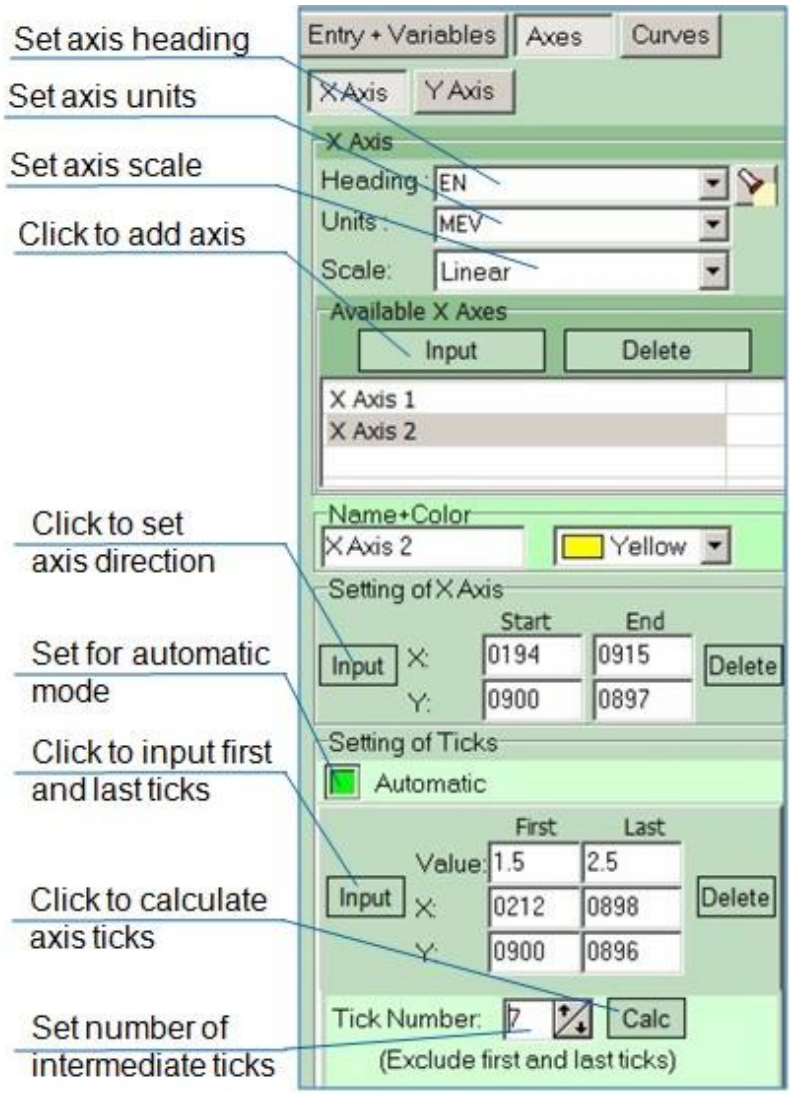

Fig. 4 − Setting of axes parameters

The next step is curves digitization itself (see. Fig.5). X and Y axes specified earlier should be associated with each new input curve. The shape and color of the markers can be selected for each curve. The input of points with symmetric and asymmetric errors along both axes is provided. Possible number of curves on the plot is also unlimited. For digitization convenience and precision enhancement there is recommended to use a lens with the fixed magnification factor.

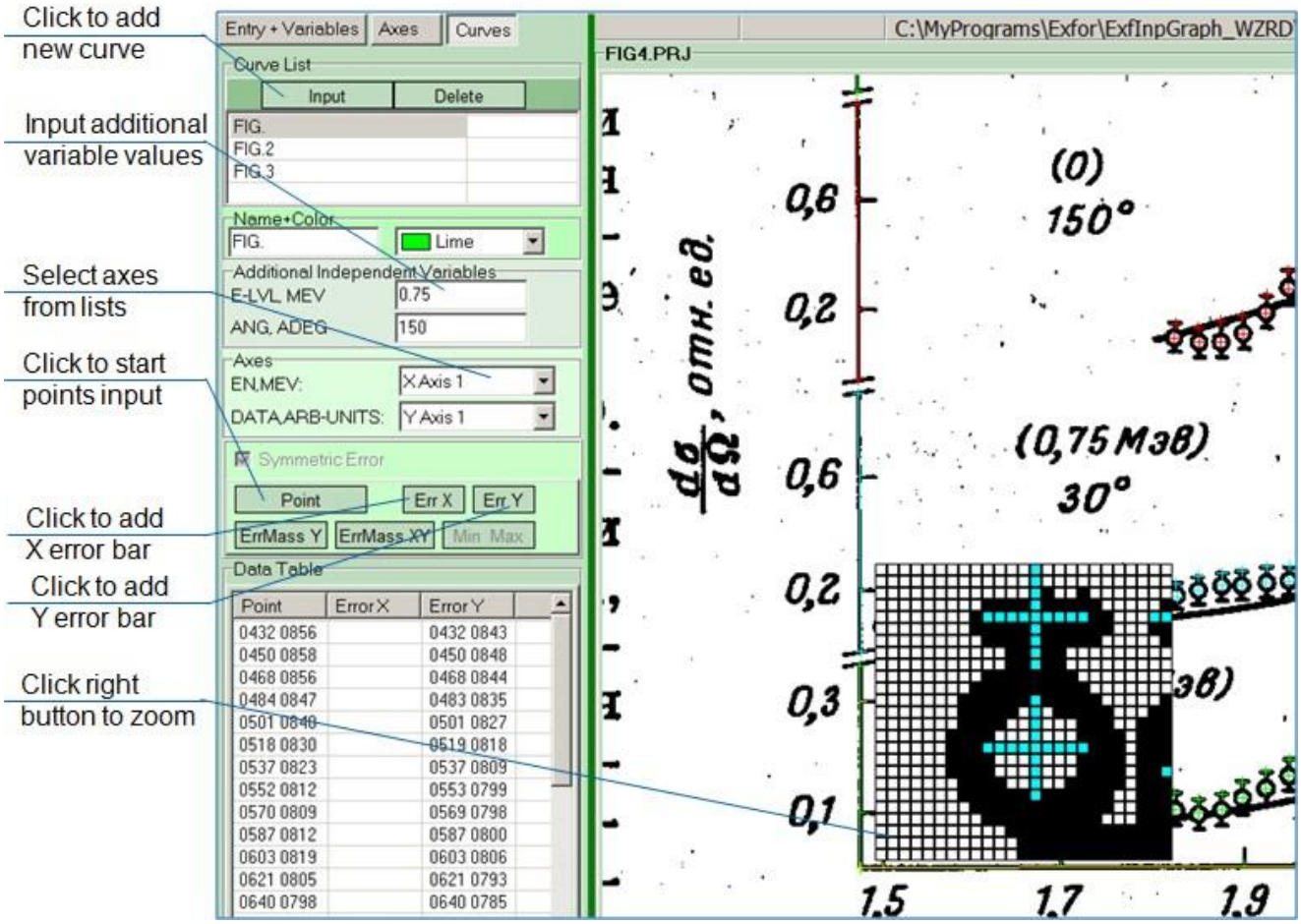

Fig. 5 – Curve digitizing

After the completion of curves digitizing the input data are compiled in the EXFOR format. The results of compilation are displayed in a special dialogue window presented in Fig. 6. The possibility of numeric data processing according to the EXFOR rules is provided in this very window [9].

For this purpose the spreadsheet mode (the DataTable mode) was introduced into the InpGraph program. It implements the following possibilities: input and editing of numeric data; setting of numeric data precision; manipulation with table rows and columns; calculations; sorting of rows into three columns; graphic presentation of numeric data; verification of data correctness; numeric data export and import in text format, Microsoft Word and Microsoft Excel formats.

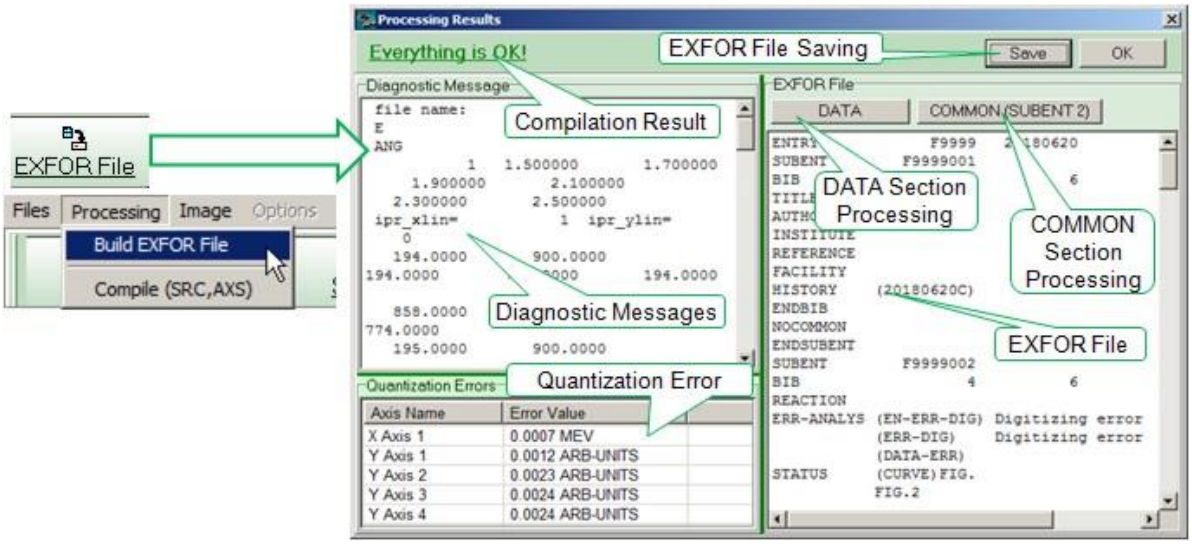

Fig. 6 – Window compilation results in the EXFOR format

Finally, the InpGraph program forms completely the sections of numeric data: DATA SECTION and COMMON SECTION − according to the EXFOR rules.

In the InpGraph program two types of digitizing errors: systematic error and quantization error are identified.

The systematic digitizing error is calculated as a mean-square (standard) deviation of the ticks introduced along X- and Y- axes in the coordinate system of graphic representation. The quantization error [10] is calculated as a half of an image pixel size expressed in physical coordinates. For original image of poor quality the quantization error can be essentially higher than the systematic error. In this case it should not be ignored.

Thus, the digitizer InpGraph has the following advantages:

− taking into account of axes ticks nonlinearity;

− taking into account of axes nonorthogonality;

− calculation of systematic error and quantization error at converting image coordinates to physical coordinates;

− possibility of experienced users to correct the digitization process by making changes in the service files;

− easy-to-use interface;

− possibility to interrupt the digitization process and resume it at any stage, make corrections and changes in the information digitized.

InpGraph may serve as an example of fruitful cooperation between the members of NRDC. They are active participants in the InpGraph testing, sending feedbacks and proposals. Our plans consist in further development and updating of the InpGraph program.

References

1. Ieruslimov A.P. and others System of Plots Digitizing (SPD) // Kratkie Soobshcheniya OIYaI, №6 [86]-97, P. 97-102.
2. https://www-nds.iaea.org/exfor.htm, https://www-nds.iaea.org/exfor/x4stat/exfor_stat.htm
3. Pikulina G.N., Taova S.M., Dunaeva S.A. // Vestn. Natz. Issl. Yader. Univer. “MIFI”. 2013. V. 2. № 1. P. 125. [in Russian].
4. http://www.frantz.fi/software/g3data.php
5. http://www.getdata-graph-digitizer.com
6. http://www.BlueLeafSoftware.com
7. http://www.nuclear.nsdc.cn/gdgraph/
8. http://www.jcprg.org/gsys/
9. Summary report of the consultants’ meeting on benchmarking of digitizing software. 12−14 November, 2012. IAEA, Vienna, Austria. INDC(NDS)-0629, Distr. G+NC.
10. Michael W. Frith “Measuring error in manually digitized maps”, Thesis, University of Regina, Redlands, California, 1997.